Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/6700--01-8537

Modeling Viral Epidemiology in Connected Networks

WILLIAM M. SPEARS

Artificial Intelligence Center
Information Technology Division

LORA BILLINGS
IRA B. SCHWARTZ

Special Project for Nonlinear Science
Plasma Physics Division

March 26, 2001

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 26, 2001	Interim 1 Oct. 00 - 1 Feb. 01

4. TITLE AND SUBTITLE

Modeling Viral Epidemiology in Connected Networks

5. FUNDING NUMBERS

PE - 62232N

6. AUTHOR(S)

William M. Spears,[1] Lora Billings,[2,3] and Ira B. Schwartz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/6700--01-8537

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington ,VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

[1]Naval Research Laboratory, Artificial Intelligence Center, Code 5515
[2]ONR/ASEE Postdoc., Code 6700.3
[3]Naval Research Laboratory, Special Project in Nonlinear Science, Code 6700.3

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

We derive some Markovian and differential equation models of viral epidemiology on connected networks. We examine the stability properties of endemic states for models based on contact rates derived from probabilistic concepts of connectivity. In particular, boundaries that delineate the onset of viral propagation in a connected network are derived and analyzed.

14. SUBJECT TERMS

Population dynamics
Virus epidemiology
Nonlinear dynamics
Computer virus

15. NUMBER OF PAGES

16

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18
298-102

MODELING VIRAL EPIDEMIOLOGY IN CONNECTED NETWORKS*

WILLIAM M. SPEARS[1,†], LORA BILLINGS[2,‡], AND IRA B. SCHWARTZ[2]

[1]AI CENTER - CODE 5515,

[2]SPECIAL PROJECT IN NONLINEAR SCIENCE - CODE 6700.3

PLASMA PHYSICS DIVISION

NAVAL RESEARCH LABORATORY, WASHINGTON, DC 20375

February 14, 2001

Abstract. We derive some Markovian and differential equation models of viral epidemiology on connected networks. We examine the stability properties of endemic states for models based on contact rates derived from probabilistic concepts of connectivity. In particular, boundaries which delineate the onset of viral propagation in a connected network are derived and analyzed.

Key words. viral epidemiology, Markov chains, transient behavior.

1. Introduction. We are becoming increasingly dependent on large interconnected networks for the control of our resources, such as the Internet, communications networks, and power grids. The advantage of these networks is the ability to route resources in a reasonably optimal fashion. However, their interconnectivity, coupled with the lack of global view of what is happening in these networks, can lead to tremendous problems in network reliability. For example, small local failures can easily propagate to entire networks, causing loss of service and corruption of data. Also, deliberate attacks, such as viral attacks, can easily cause widespread havoc. The purpose of this paper is to outline a broad framework for describing the spread of viruses. The framework is designed to be as general as possible, with restrictive assumptions only introduced when needed. The focus will be on two complementary analysis techniques – discrete-time Markov chain models and continuous-time differential equation models. We give a brief overview of Markov chains first.

2. Overview of Markov Chains. A discrete-time Markov chain is a dynamical system composed of S discrete states. At each time step, the Markov chain can change states. Let X_t be the random variable for the Markov chain, which can take on any of the S states at time t. The system is described by an $S \times S$ matrix Q, which gives the probability of transitioning from one state a at time $t-1$ to state b at time t, where a component of Q is given by:

$$Q(a,b) \equiv p_{a,b} \equiv P(X_t = b \mid X_{t-1} = a)$$

The $p_{a,b}$ values define the "one-step" probability transition matrix Q, since Q describes the probability of transitioning from state to state in one time step. The transient behavior of the system is obtained from the "n-step" probability transition values, which are obtained from the nth power of Q:

$$Q^n(a,b) \equiv p_{a,b}{}^{(n)} \equiv P(X_t = b \mid X_{t-n} = a)$$

It is also possible to compute conditional probabilities over a set of states. Define a predicate $Pred_B$ and the set B of states that make $Pred_B$ true. Then the probability

*This work was supported by ONR.

†Phone: 202-767-9006, FAX: 202-767-3172, spears@aic.nrl.navy.mil

‡Supported by ONR/ASEE postdoc

Manuscript approved February 20, 2001.

that the system will be in one of the states of B at time t, given it is in state a at time $t-n$ is:

$$p_{a,B}{}^{(n)} \equiv P(X_t \in B \mid X_{t-n} = a) = \sum_{b \in B} p_{a,b}{}^{(n)}$$

Analysis of Markov chain behavior can be considered to be "instantaneous" or "cumulative." Instantaneous behavior refers to events that occur *at* a particular time. An example would be whether the system is at a particular state at that time. Cumulative behavior refers to events that have occurred *by* a certain time.

2.1. Instantaneous Transient Behavior at Time n. The probability of being in some state b at time n is given by simply considering the probability of each possible n-step transition, appropriately weighted by the a priori probabilities:

$$p_b{}^{(n)} \equiv P(X_n = b) = \sum_a p_a{}^{(0)} \, p_{a,b}^{(n)}$$

where the a priori probability of a system being in state a at time 0 is denoted as $p_a{}^{(0)}$.

The probability that the system is in one of the states of B at time n is:

$$p_B{}^{(n)} = \sum_{b \in B} p_b{}^{(n)} \tag{2.1}$$

Finally, it is possible to compute the probability that the system will transition from one set of states to another set of states. Let $Pred_A$ be another predicate over the states, and denote A to be the set of states that make $Pred_A$ true. Then the probability that the system will be in one of the states of B at time t, given that it is in one of the states of A at time $t-n$, is:

$$p_{A,B}{}^{(n)} \equiv \tag{2.2}$$
$$P(X_t \in B \mid X_{t-n} \in A) = \frac{\sum_{a \in A} p_a{}^{(t-n)} \, p_{a,B}{}^{(n)}}{p_A{}^{(t-n)}} = \frac{\sum_{a \in A} p_a{}^{(t-n)} \, p_{a,B}{}^{(n)}}{\sum_{a \in A} p_a{}^{(t-n)}}$$

which involves a renormalization over the states indexed by A. Since Eq. 2.2 describes how a system transitions from a group of states to another group of states, this raises the intriguing notion that a system with a large number of states might be simplified (or aggregated) into a system with a smaller number of groups of states. Unfortunately, this is hard to do in general since Eq. 2.2 is a time-dependent equation (thus making the aggregated Markov chain non-stationary), however some progress has been made in this area [3]. The nice feature of this formalization is that *any* predicate over the states can be used.

It is possible to generalize further to arbitrary functions f over the states and compute, for example, the expected value of that function, at time n:

$$E[f]^{(n)} = \sum_b p_b{}^{(n)} \, f(b) \tag{2.3}$$

2.2. Cumulative Transient Behavior. Another common computation involving Markov chains is referred to as the "mean first passage times" for going from state a to state b (for a nice discussion of this, see [4]). This refers to the length of time that

it takes (on the average) to reach state b for the first time, given that the process is currently in state a. Answering such questions involves solving the set of simultaneous equations:

$$m_{a,b} \equiv p_{a,b} + \sum_{k \neq b} p_{a,k} \, (1 + m_{k,b}) \tag{2.4}$$

where $m_{a,b}$ denotes the mean first passage time from state a to state b. To understand the equation, consider transitioning from state a to b in one move. This occurs with probability $p_{a,b}$ and requires only one step. However, suppose the system transitions from state a to state k, where k is not equal to b. This occurs with probability $p_{a,k}$ and requires one step. However, there now remain $m_{k,b}$ steps to state b.

As before, if there is interest in a set B of states, it is possible to compute the mean first passage time for the system to first enter that set of states, given that it is currently outside that set:

$$m_{a,B} \equiv \sum_{b \in B} p_{a,b} + \sum_{k \notin B} p_{a,k} \, (1 + m_{k,B})$$

where $m_{a,B}$ denotes the mean first passage time from state a to any of the states in set B, and a is not in B. This is very similar to Eq. 2.4, with the exception that the probability of entering state B in one step is simply the sum of the probabilities of entering each state within B.

Once this system of simultaneous equations is solved, it is possible to calculate the "expected waiting time" to reach a state in B, given a random initial state, via:

$$EWT_B = \sum_{a \in B} p_a{}^{(0)} \, 0 + \sum_{a \notin B} p_a{}^{(0)} \, m_{a,B}$$

There are two parts to this equation. The first part reflects the possibility that a random initial population is in state B, and hence has a zero waiting time. The second part reflects the mean passage time from initial populations not in B, to a state in B. Clearly this simplifies to:

$$EWT_B = \sum_{a \notin B} p_a{}^{(0)} \, m_{a,B}$$

This holds for any set of states B, and thus it can be used to provide expected waiting times for a variety of events.

3. Markov Models of Viral Spread. In this paper we are concerned with the spread of viruses throughout a network. We assume that the network consists of N nodes. Each node can be in one of four *medical conditions*: Susceptible (S), Exposed (E), Infected (I), and Cured (C). It is often customary to also refer to these conditions as "states", but that would cause some confusion with the Markov definition of state. In this paper the number of nodes in the various medical conditions will define a Markov state. In general, if there are m medical conditions and N nodes, the number of Markov states S can be shown to be:

$$S = \begin{pmatrix} N + m - 1 \\ m - 1 \end{pmatrix}$$

3.1. The Two-Condition I-C Case. We first start by ignoring the Susceptible and Exposed conditions, and assume that the nodes can either be infected or not infected (cured). Infected nodes can remain infected or become cured. Likewise, cured nodes can remain cured or become infected. This is the simplest case to analyze. Figure 3.1 illustrates the transitions between these two medical conditions, with their associated probabilities.

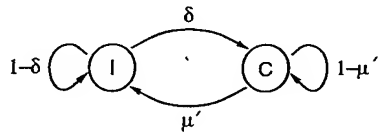

FIG. 3.1. *The transition diagram showing how a node can change medical conditions.*

Let I be the number of infected nodes and C be the number of cured nodes. Since each node must be infected or cured at any given time, $I + C = N$. The Markovian state is defined by the pair (I, C) and we want to compute the probability of transitioning from state $a = (I, C)$ to state $b = (I', C')$:

$$p_{a,b} \equiv P(X_t = (I', C') \mid X_{t-1} = (I, C))$$

Such a transition means that $I' - I$ *more* cured nodes are infected than infected nodes are cured. Note that if $I' < I$ then the previous sentence is still mathematically and grammatically correct, but it is perhaps more understandable to say that $I - I'$ *less* cured nodes are infected than infected nodes are cured.

Let x be the number of infected nodes that are cured. Then since there are I infected nodes, $I - x$ of them are *not* cured. If we use δ to denote the probability of curing an infected node, then the probability of curing x of the infected nodes is simply:

$$\binom{I}{x} (\delta)^x (1-\delta)^{I-x}$$

Also, since $I' - I$ *more* cured nodes are infected than infected nodes are cured, we know that $x + I' - I$ of the cured nodes must become infected. Since there are C cured nodes then $C - (x + I' - I)$ of the cured nodes are *not* infected. If we use μ' to denote the probability of infecting a cured node, then the probability of infecting $x + I' - I$ of the cured nodes is simply:

$$\binom{C}{x + I' - I} (\mu')^{x+I'-I} (1-\mu')^{C-(x+I'-I)}$$

We can put this all together by summing over all possible values of x:

$$\begin{aligned} p_{a,b} = \sum_x & \binom{I}{x} (\delta)^x (1-\delta)^{I-x} \times \\ & \binom{C}{x + I' - I} (\mu')^{x+I'-I} (1-\mu')^{C-(x+I'-I)} \end{aligned} \tag{3.1}$$

It now only remains to bound x correctly, which can be achieved by examining the combinatorials. The first combinatorial implies that $0 \leq x \leq I$. The second implies

that $I - I' \leq x \leq C - I' + I$. In order for all of these constraints to be held we get:

$$p_{a,b} = \sum_{x=max\{0,I-I'\}}^{min\{I,C-I'+I\}} \binom{I}{x} (\delta)^x (1-\delta)^{I-x} \times \binom{C}{x+I'-I} (\mu')^{x+I'-I} (1-\mu')^{C-(x+I'-I)} \tag{3.2}$$

Since $C = N - I$, this equation can be slightly simplified, but we do not do that here, in order to provide a smooth transition to the three-condition case. It is also useful to note that the number of Markov states for the two-medical condition case is simply:

$$\mathcal{S} = \binom{N+2-1}{2-1} = N+1$$

3.2. The Three-Condition S-I-C Case. We now add in the Susceptible condition. Susceptible nodes can remain susceptible or become infected. Infected nodes can remain infected or become cured. Cured nodes can remain cured or become susceptible. Figure 3.2 illustrates the transitions between these three medical conditions, with their associated probabilities.

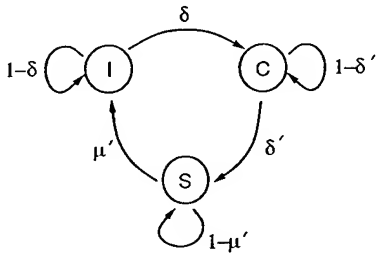

FIG. 3.2. *The transition diagram showing how a node can change medical conditions.*

Again, let I be the number of infected nodes and C be the number of cured nodes. Let S be the number of susceptible nodes. Since each node must be susceptible, infected or cured at any given time, $S + I + C = N$. The Markovian state is defined by the triple (S, I, C) and we want to compute the probability of transitioning from state $a = (S, I, C)$ to state $b = (S', I', C')$:

$$p_{a,b} \equiv P(X_t = (S', I', C') \mid X_{t-1} = (S, I, C))$$

Such a transition means that $I' - I$ *more* susceptible nodes are infected than infected nodes are cured. It also means that $S' - S$ *more* cured nodes become susceptible than susceptible nodes become infected.

Let x be the number of infected nodes that are cured. Then since there are I infected nodes, $I - x$ of them are *not* cured. If we use δ to denote the probability of curing an infected node, then the probability of curing x of the infected nodes is again:

$$\binom{I}{x} (\delta)^x (1-\delta)^{I-x}$$

Also, since $I' - I$ *more* susceptible nodes are infected than infected nodes are cured, we know that $x + I' - I$ of the susceptible nodes must become infected. Since there are S susceptible nodes then $S - (x + I' - I)$ of the susceptible nodes are *not* infected. If we use μ' to denote the probability of infecting a susceptible node, then the probability of infecting $x + I' - I$ of the susceptible nodes is simply:

$$\binom{S}{x + I' - I} (\mu')^{x+I'-I}(1-\mu')^{S-(x+I'-I)}$$

Finally, since $S' - S$ *more* cured nodes become susceptible than susceptible nodes become infected, we know that $x + I' - I + S' - S$ of the cured nodes must become susceptible. Since there are C cured nodes then $C - (x + I' - I + S' - S)$ of the cured nodes are *not* made susceptible. If we use δ' to denote the probability of making a cured node susceptible, then the probability of making $x + I' - I + S' - S$ of the cured nodes susceptible is simply:

$$\binom{C}{x + I' - I + S' - S} (\delta')^{x+I'-I+S'-S}(1-\delta')^{C-(x+I'-I+S'-S)}$$

Putting this all together we get:

$$\begin{aligned} p_{a,b} = \sum_{x=max\{0,I-I',I-I'+S-S'\}}^{min\{I,S-I'+I,C-I'+I-S'+S\}} & \binom{I}{x} (\delta)^x(1-\delta)^{I-x} \times \\ & \binom{S}{x + I' - I} (\mu')^{x+I'-I}(1-\mu')^{S-(x+I'-I)} \times \\ & \binom{C}{x + I' - I + S' - S} (\delta')^{x+I'-I+S'-S}(1-\delta')^{C-(x+I'-I+S'-S)} \end{aligned} \tag{3.3}$$

For the three-medical condition case the number of Markov states is given by:

$$\mathcal{S} = \binom{N+3-1}{3-1} = \frac{(N+2)(N+1)}{2}$$

3.3. The Four-Condition S-E-I-C Case. It is possible to add in the Exposed condition. A susceptible node becomes exposed before it becomes infected. Figure 3.3 illustrates the transitions between these four medical conditions, with their associated probabilities.

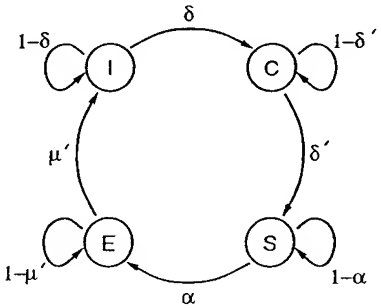

FIG. 3.3. *The transition diagram showing how a node can change medical conditions.*

The Markovian state is defined by the tuple (S, E, I, C) and we want to compute the probability of transitioning from state $a = (S, E, I, C)$ to state $b = (S', E', I', C')$:

$$p_{a,b} \equiv P(X_t = (S', E', I', C') \mid X_{t-1} = (S, E, I, C))$$

Such a transition means that $I' - I$ *more* exposed nodes are infected than infected nodes are cured. It also means that $E' - E$ *more* susceptible nodes become exposed than exposed nodes become infected. Finally, $S' - S$ *more* cured nodes become susceptible than susceptible nodes become exposed.

It is easy to produce $p_{a,b}$ by extending Equation 3.3:

$$\begin{aligned} p_{a,b} = & \sum_{x=max\{0,I-I',I-I'+E-E',I-I'+E-E'+S-S'\}}^{min\{I,E-I'+I,S-I'+I-E'+E,C-I'+I-E'+E-S'+S\}} \binom{I}{x} (\delta)^x (1-\delta)^{I-x} \times \\ & \binom{E}{x+I'-I} (\mu')^{x+I'-I} (1-\mu')^{E-(x+I'-I)} \times \\ & \binom{S}{x+I'-I+E'-E} (\alpha)^{x+I'-I+E'-E} (1-\alpha)^{S-(x+I'-I+E'-E)} \times \\ & \binom{C}{x+I'-I+E'-E+S'-S} (\delta')^{x+I'-I+E'-E+S'-S} (1-\delta')^{C-(x+I'-I+E'-E+S'-S)} \end{aligned} \tag{3.4}$$

For the four-medical condition case the number of Markov states is given by:

$$\mathcal{S} = \binom{N+4-1}{4-1}$$

3.4. Discussion. Given the above Markov models, we can use them to determine various quantities of interest. For example, the probability distribution of infected nodes at time n can be obtained, as well as the probability of extinction. The expected number of infected nodes at time n is easily computed. Finally, the mean passage time until viral extinction is easily obtained.

Given these preliminary models, it should be clear at this point that they can easily be expanded to include m medical conditions, assuming that the transitions between medical conditions occur in a ring (as is true for the models examined thus far). Another important point is that the parameters δ, μ', α, and δ' need *not* be constants – they in fact can be *functions* that depend on the state that the process is in. This level of generality can provide for nonlinear dynamics of the system, as we will see in the following discussion concerning differential equation models of viral spread.

4. Differential Equation Models of Viral Spread. We can now provide analogous differential equation models.

4.1. The Two-Condition I-C Case. If we let $i = I/N$ and $c = C/N$ be the *proportion* of infected and cured nodes (respectively) then:

$$\begin{aligned} \frac{di}{dt} &= c\mu' - i\delta \\ \frac{dc}{dt} &= i\delta - c\mu' \end{aligned} \tag{4.1}$$

Since $i + c = 1$ this can be simplified to:

$$\frac{di}{dt} = c\mu' - i\delta = (1 - i)\mu' - i\delta \tag{4.2}$$

where μ' is the probability of infecting a cured node and δ is the probability of curing an infected node.

4.2. The Three-Condition S-I-C Case. If we add in the Susceptible condition then the system can be described as follows:

$$\begin{aligned} \frac{ds}{dt} &= c\delta' - s\mu' \\ \frac{di}{dt} &= s\mu' - i\delta \\ \frac{dc}{dt} &= i\delta - c\delta' \end{aligned} \tag{4.3}$$

where $s = S/N$ is the proportion of susceptible nodes. In this case μ' is the probability of infecting a susceptible node, δ is the probability of curing an infected node, and δ' is the probability of making a cured node susceptible. This could be reduced to two equations if desired, assuming a constraint of $s + i + c = 1$.

4.3. The Four-Condition S-E-I-C Case. If we add in the Exposed condition then the system can be described as follows:

$$\begin{aligned} \frac{ds}{dt} &= c\delta' - s\alpha \\ \frac{de}{dt} &= s\alpha - e\mu' \\ \frac{di}{dt} &= e\mu' - i\delta \\ \frac{dc}{dt} &= i\delta - c\delta' \end{aligned} \tag{4.4}$$

where $e = E/N$ is the proportion of exposed nodes. In this case μ' is the probability of infecting an exposed node, δ is the probability of curing an infected node, δ' is the probability of making a cured node susceptible, and α is the probability of making a susceptible node exposed. This could be reduced to three equations if desired, assuming a constraint of $s + e + i + c = 1$.

4.4. Discussion. Again, it should now be very clear that the differential equation models can easily be expanded to include m medical conditions, assuming that the transitions between medical conditions occur in a ring (as is true for the models examined thus far). Also, it is important to reiterate that the parameters δ, μ', α, and δ' need *not* be constants, but may be functions that depend on the state of the process (e.g., the proportion of infected nodes i). A concrete example is shown in the following section.

5. Random Directed Graphs and Measures of Network Connectivity. The nice aspect of the above framework is that we have not had to yet define μ', α, δ, and δ'. In this paper we will assume that δ is a constant that reflects an infected node's ability to cure itself (the strength of the immune system or the ability of the sysop to rid the node of the virus). In the three-condition case, δ' is a constant

that reflects a cured node's frequency of opening itself back up for infection (by weakening the immune system through malnutrition or becoming lazy as a sysop). In the four-condition case, α is a constant that reflects the probability of making a susceptible node exposed, and could be analogous to the likelihood of going to the mall or connecting a computer to the net.[1] On the other hand, the probability that a cured/susceptible/exposed node (depending on whether we are talking about the two-, three-, or four-condition case) will become infected, μ', is largely dependent on the number of other neighboring nodes in the graph that are also infected. The computation of μ' is thus very dependent on the structure of the graph. In this paper we recalculate this quantity for the case of the random directed graphs defined in [1].

A random directed graph of N nodes is constructed by making random, independent decisions about whether to include each of the $N(N-1)$ possible directed edges that can connect two nodes. Each edge is included in the graph with probability p. Since there are N nodes, you can have at most $N-1$ neighbors pointing towards each node (one doesn't connect to oneself). In expectation any given node has $p(N-1)$ neighbors to itself. In [1] this is denoted as $\bar{b} \equiv p(N-1)$, and is a measure of the connectivity of the network. We now state the function for μ' for three separate situations.

5.1. Strong Links. In this case you can only become infected from your neighbors that are infected. The probability that an infected neighbor can transmit the infection to you is β. Thus the probability that an infected neighbor won't transmit the infection to you is $1-\beta$. Since i is the proportion of infected nodes in the network, in expectation you will have $iN\bar{b}/(N-1)$ infected neighbors (because once again the assumption is that you yourself are not infected). The probability that none will infect you is $(1-\beta)^{iN\bar{b}/(N-1)}$. Thus the probability that you will be infected by at least one infected neighbor is:

$$\mu' = 1-(1-\beta)^{\frac{iN\bar{b}}{N-1}} = 1-(1-\beta)^{iNp}$$

where once again i is the proportion of infected nodes.

If N is large, the reasonable approximation (which we will use throughout) is:

$$\mu' = 1-(1-\beta)^{i\bar{b}}$$

If y and z are sufficiently small it is well known that a reasonable approximation to $(1-y)^z$ is $1-yz$. If this approximation is taken (which is not necessarily reasonable in this situation) we obtain:

$$\mu' = \beta i\bar{b}$$

which agrees with [1]. Notice that although the systems of the two-, three-, and four-condition cases appear as linear differential equations, the effective contacts between infectives and susceptibles create a mass action term. This, in turn, makes the problem of viral spread in a network a nonlinear one.

5.2. Weak Links. In this case you can also be infected from infected nodes that are not your immediate neighbor. The probability of transmission in this case is β_w. In expectation you will have $N-\bar{b}-1$ non-immediate neighbors, of which $i(N-\bar{b}-1)$ are infected. These neighbors will *not* infect you with probability $(1-\beta_w)^{i(N-\bar{b}-1)}$.

[1] There is no reason why we can't relax these assumptions if we wish.

Putting it all together, the probability that you will be infected by at least one of your immediate or non-immediate neighbors is:

$$\mu' = 1 - (1-\beta)^{i\bar{b}}(1-\beta_w)^{i(N-\bar{b}-1)}$$

Making the same approximation used above, and throwing out further second-order terms (which again may not be reasonable to do) yields:

$$\mu' = \beta i\bar{b} + \beta_w i(N-\bar{b}-1)$$

which agrees with [1]. Very weak links in the form of mass action terms such as that derived above, can qualitatively change the dynamics in coupled populations [2].

5.3. Hierarchical Links. The final case we consider is that of hierarchical links. In this case you do not assume that you have $\bar{b}$ neighbors. Instead, neighbors are distributed around each node in L layers. The number of nodes at layer (distance) l is 2^{l-1} and there are a total of $N = 2^L$ nodes. The probability of transmission from an infected node at distance l is given by $\beta_0\tau^{l-1}$, where τ is a constant ranging from 0 to 1.

Following a similar process used above it is simple to show that the probability of being infected is given as:

$$\mu' = 1 - \prod_{l=1}^{L}(1-\beta_0\tau^{l-1})^{i2^{l-1}}$$

Making the same approximation used above, and throwing out all higher-order terms (which again may not be reasonable to do) yields:

$$\mu' = \sum_{l=1}^{L}\beta_0(2\tau)^{l-1}i$$

which agrees with [1].

6. Analysis and Comparison of the Models. Now that μ' has been computed for random-directed graphs, we are directly able to use either the Markov chain or differential equation models of virus spread on random-directed graphs. We now run simulations of the Markov models and compare the results to the differential equation models.

6.1. The Two-Condition I-C Case. The differential equation for the two-condition case is

$$\frac{di}{dt} = (1-i)(1-(1-\beta)^{i\bar{b}}) - \delta i,$$

where the constant solutions are found by solving $\frac{di}{dt} = 0$. The zero solution, $i_0 = 0$, exists for all β and δ. The other solution is called the endemic solution, i_e, which is implicitly defined on $0 \le \delta \le \min(-\bar{b}\ln(1-\beta), 1)$. The two possible behaviors of this model are divided in parameter space by the curve

$$\delta = -\bar{b}\ln(1-\beta).$$

If $\delta \ge -\bar{b}\ln(1-\beta)$, solutions tend toward the solution i_0 for all initial conditions. In other words, the virus will naturally die out. In the rest of the parameter space, the

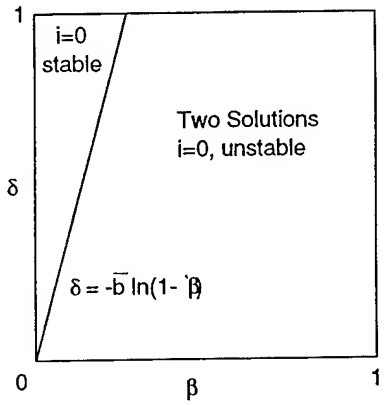

FIG. 6.1. *The stability of i_0 and i_e in the two-condition case as a function of β and δ.*

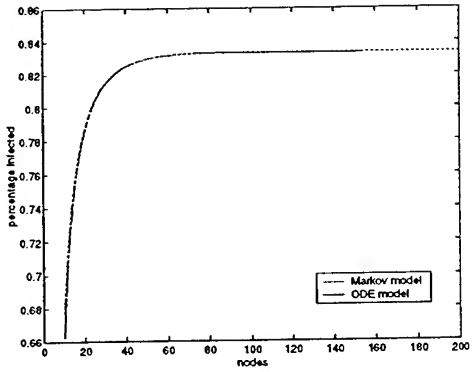

FIG. 6.2. *The agreement of the Markov and ODE models in the two-condition case as a function of the number of nodes, N. In this computation, $\beta = 0.2$, $\delta = 0.2$, $\bar{b} = 0.4(N-1)$, all of the nodes are initially infected, and the Markov model is iterated 20 time steps. The percentage of nodes infected increases exponentially to $0.8\bar{3}$ percent.*

solution i_0 is unstable and all solutions tend toward the stable, endemic solution i_e. Here, the virus will not naturally die out, but persist at a constant level. See Figure 6.1. The preferred behavior is for i_0 to be stable.

The ODE model agrees with the Markov model depending on the initial conditions. Shown in Figure 6.2, as the number of nodes is increased, the percentage of nodes infected increases exponentially. The region where i_0 is stable is also predicted by the Markov model, with the division matching the curve $\delta = -\bar{b}\ln(1-\beta)$ as shown in Figure 6.3.

6.2. The Three-Condition S-I-C Case. In two variables, assuming $s+i+c = 1$, the differential equation for the three-condition case is

$$\begin{aligned} \frac{ds}{dt} &= \delta'(1-i-s) - (1-i)(1-(1-\beta)^{i\bar{b}}) \\ \frac{di}{dt} &= (1-(1-\beta)^{i\bar{b}})s - \delta i. \end{aligned} \tag{6.1}$$

The zero fixed point is $\mathrm{p}_0 = (0,1)$. The eigenvalues of the linearization about p_0 are $\lambda_1 = -\delta'$ and $\lambda_2 = -\delta - \bar{b}\ln(1-\beta)$. Since $0 \le \delta' \le 1$, then the eigenvalue $\lambda_1 \le 0$.

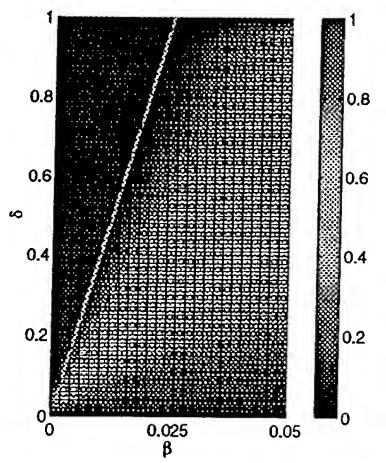

FIG. 6.3. *For the two-condition case, the stable solution i predicted by the Markov model after 20 time steps as a function of β and δ. Here, $N = 100$ and $\bar{b} = 0.4(N-1)$. Superimposed in green is the curve $\delta = -\bar{b}\ln(1-\beta)$ from the ODE model dividing where i_0 and i_e are stable.*

The other eigenvalue, λ_2, is negative for $1 \geq \delta > -\bar{b}\ln(1-\beta)$, where $\mathbf{p}_0$ a stable fixed point. Note that this is the same curve as in the two-condition case. The associated eigenvectors are

$$\begin{aligned} \mathbf{e}_1 &= (0,1) \\ \mathbf{e}_2 &= \left(\frac{\delta - \delta' + \bar{b}\ln(1-\beta)}{\delta' - \bar{b}\ln(1-\beta)}, 1\right). \end{aligned} \tag{6.2}$$

In the rest of parameter space, $\mathbf{p}_0$ is an unstable saddle fixed point. That is because $0 \leq \delta \leq \min(-\bar{b}\ln(1-\beta), 1)$, making the eigenvalue λ_2 non-negative.

The endemic fixed point, $\mathbf{p}_e = (i_e, s_e)$, is defined implicitly:

$$\begin{aligned} 0 &= (1-i_e)\delta' - i_e\delta - \frac{i_e\delta\delta'}{1-(1-\beta)^{i_e\bar{b}}} \\ s_e &= \frac{i_e\delta}{1-(1-\beta)^{i_e\bar{b}}} \end{aligned} \tag{6.3}$$

For an example solution, see Figure 6.4. The eigenvalues for this fixed point are complex conjugates with norm less than 1, or stable, for parameters where the zero solution, $\mathbf{p}_0$ is unstable. The expression $\delta = -\bar{b}\ln(1-\beta)$ defines a surface in parameter space (β, δ, δ') on which $\mathbf{p}_0$ has neutral stability. This surface divides where the virus persists from where the virus dies out.

It would now be possible to run the three-condition case simulation for the Markov model and compare the results to the ODE model, although we leave that for the future.

7. Conclusions. We have derived a probabilistic framework that provides a more accurate representation of how viruses spread throughout a network. In particular, the framework lays down the building blocks on which to incorporate more realistic parameters, such as measures of connectivity, effective contact rates, and different network architectures.

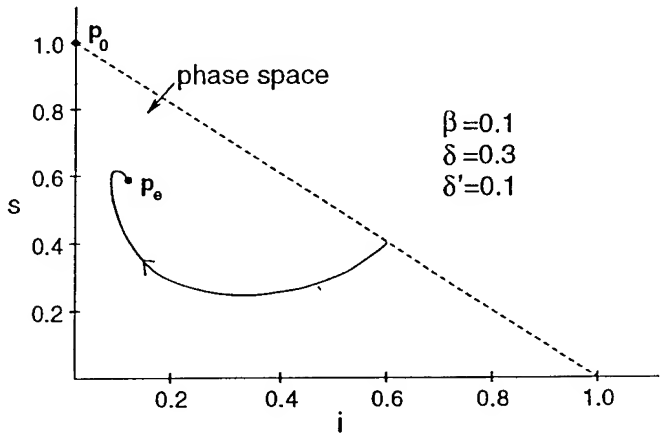

FIG. 6.4. *Sample solution in the three-condition case for initial condition* $(i, s, c) = (0.6, 0.4, 0)$, $\beta = 0.1$, $\delta = 0.3$, $\delta' = 0.1$, *and* $\bar{b} = 5$.

The framework is expressed in terms of Markov chain and differential equation models. Although the Markov chain models are inherently more accurate, they are also computationally more intractable. The differential equation models provide a viable alternative approach. Simple dynamic models expressing the state variables have been derived for large populations. The effective contact rates have been expressed in terms of network connectivity measures, and generate a viral spread that is the result of nonlinear mass action.

Analysis of the models shows that for an endemic state to persist, the trivial solution of all susceptibles and no infectives must be unstable. A curve defining neutral stability between die out and endemicity has been derived and computed based on a stability exchange between the two states. Such curves provide a metric in which to measure the spread of a virus in a network, and may provide a future parameter in which to control the spread without sacrificing much network information flow.

Acknowledgments. This work was supported by the Office of Naval Research.

REFERENCES

[1] J. KEPHART AND S. WHITE, *Directed-graph epidemiological models of computer viruses*, in Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy, 1991, pp. 343–359.

[2] I. B. SCHWARTZ, *Small amplitude, long period outbreaks in seasonally driven epidemics*, J. Math. Biol., 30 (1992), pp. 473–492.

[3] W. SPEARS, *A compression algorithm for probability transition matrices*, SIAM Matrix Analysis and Applications, 20 (1998), pp. 60–77.

[4] W. WINSTON, *Operations research: Applications and Algorithms*, PWS-Kent, Boston, MA, 1991.